Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the *HNF4A* region

The UK IBD Genetics Consortium
The Wellcome Trust Case Control Consortium 2

Supplementary Tables

SNP	Chr	Position	P_{scan}	P_{repl}
rs12568930*	1p36.12	22574818	1.4×10^{-7}	0.05
rs17019600	2p12	80769245	5.1×10^{-5}	0.7
rs6437358	2q37.3	241245823	5.9×10^{-7}	0.13
rs4543390	6q24.3	146865778	1.1×10^{-6}	0.3
rs886774	7q31.1	107282670	4.8×10^{-7}	0.005
rs7020238	9p24.1	7199284	2.7×10^{-5}	0.99
rs10781500	9q34.3	138389159	1.3×10^{-5}	0.038
rs12271425	11p15.2	13882135	2.1×10^{-5}	FAIL
rs3897233	13q12.13	26440276	1.2×10^{-6}	0.28
rs9548988	13q13.3	39403510	5.0×10^{-6}	0.0061
rs17104722	14q24.3	76207863	3.9×10^{-6}	0.22
rs1428103	16p12.3	17931478	6.0×10^{-6}	0.32
rs2764742	16p12.3	19938333	6.8×10^{-7}	0.13
rs1728785	16q22.1	67148731	1.8×10^{-5}	0.0004
rs13337840	16q24.1	85384840	9.0×10^{-7}	0.34
rs6017342	20q13.12	42498442	3.2×10^{-13}	7.1×10^{-6}

Supplementary Table 1: Results for all SNPs attempted in the replication experiment. *rs12568930 is a perfect proxy for the most associated SNP in the GWA, rs7524102.

SNP	Chr	Gene of interest	P_{crohn}
rs11209026	1p31.3	*IL23R*	6.38×10^{-34}
rs7524102	1p36.12		0.095
rs6426833	1p36.13	*OTUD3/PLA2GE*	0.397
rs7511649	1q21.2	*ECM1*	0.360
rs3024493	1q32.1	*IL10*	Not in GWAS
rs7554511	1q32.1	*KIF21B*	2.68×10^{-6}
rs12612347	2q35	*ARPC2*	0.076
rs9858542	3p21.31	*MST1*	1.89×10^{-7}
rs10021288	4q27	*IL2/21*	Not in GWAS
rs1368438	5q33.3	*IL12B*	0.098
rs9268877	6p21.32	MHC	0.008
rs6908425	6p22.3	*CDKAL1*	2.53×10^{-7}
rs12529198	6p25.1	*LYRM4*	7.08×10^{-7}
rs886774	7q31.1	*LAMB1*	0.280
rs10974914	9p24.1	*JAK2*	Not in GWAS
rs10781500	9q34.3	*CARD9*	Not in GWAS
rs17582416	10p11.21	*CCNY*	8.48×10^{-6}
rs10995271	10q21.2		1.90×10^{-11}
rs6584283	10q24.2	*NKX2-3*	3.04×10^{-10}
rs12815372	12q15	*IL26*	Not in GWAS
rs9548988	13q13.3		0.025
rs916977	15q13.1	*HERC2*	0.084
rs1728785	16q22.1	*CDH1*	0.549
rs744166	17q21.2	*STAT3*	5.94×10^{-6}
rs2542151	18p11.21	*PTPN2*	1.19×10^{-11}
rs6017342	20q13.12	*HNF4A*	0.768
rs311497	20q13.33	*TNFRSF6B*	Not in GWAS
rs2094871	21q22.2	*PSMG1*	Not in GWAS

Supplementary Table 2: Evidence for association with Crohn's disease meta-analysis[6] for UC loci described in this paper.

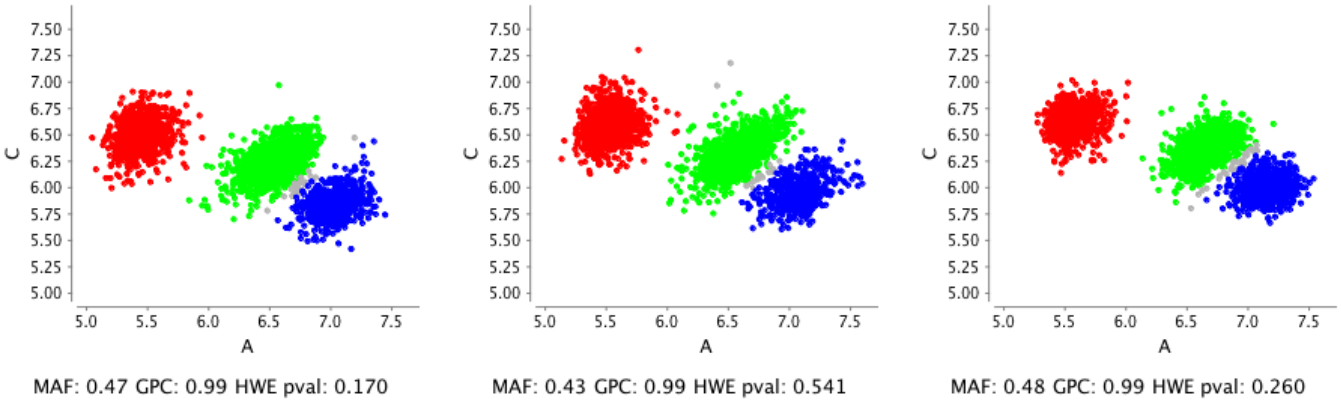
	QC	Ancestry	Relatedness	Gender	Total
UC	226	85	84	60	413
1958BC	189	47	14	5	255
UKBS	185	38	81	32	312

Supplementary Table 3: Breakdown of the number of individuals removed from each collection. QC: heterozygosity & missingness; Ancestry: HapMap PCA; Relatedness: inferred IBD >5%; Gender: conflict with manifest or uncertain; Total: unique individuals.

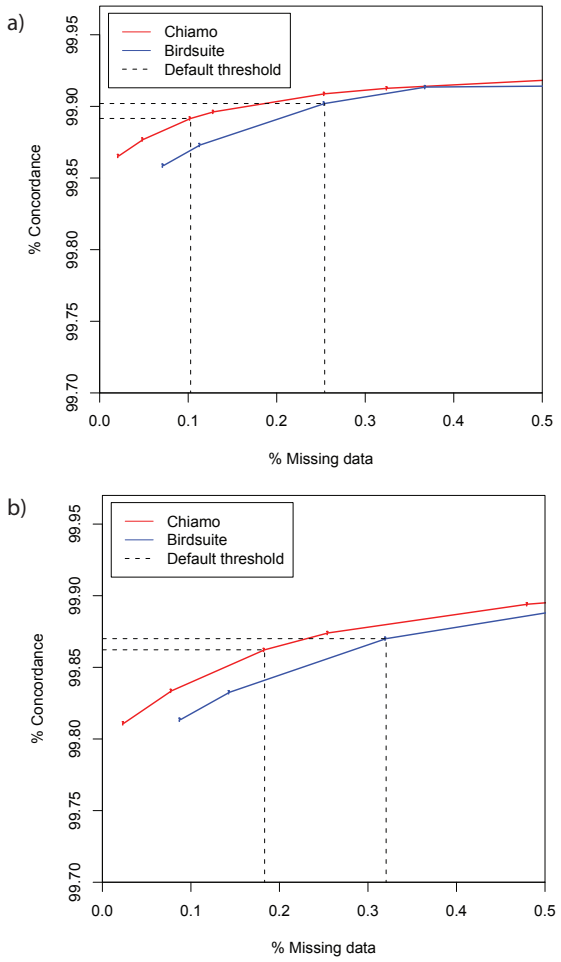
	GWAS	Replication
Cambridge	715	239
Dundee	-	189
Edinburgh	258	282
Exeter	-	338
London	333	78
Manchester	-	189
Newcastle	215	326
Oxford	394	247
Sheffield	446	60
Torbay	-	373
Total	2361	2321

Supplementary Table 4: Centre of origin for cases used in both GWAS and replication collections.

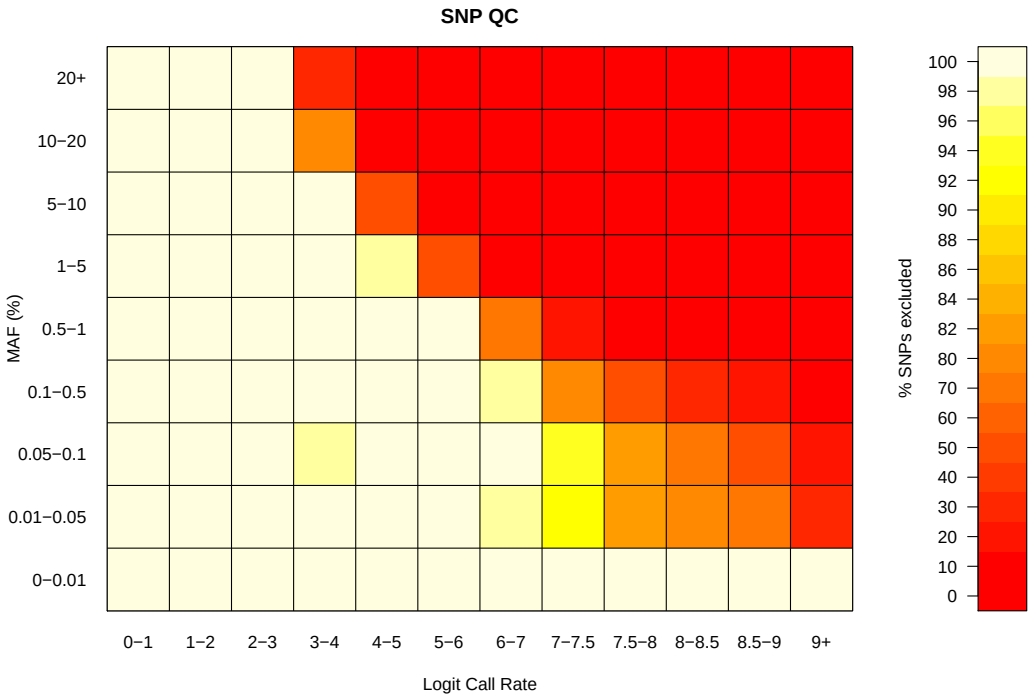
Supplementary Figures



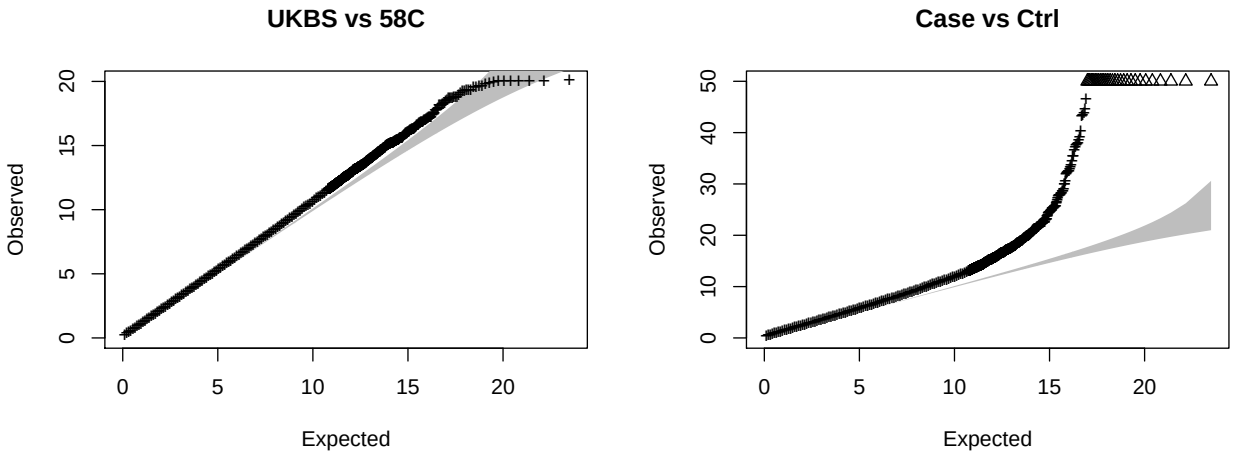
Supplementary Figure 1: Genotype cluster plots for SNP rs6017342 in each of the three collections (UKBS, UC, 1958BC). Minor allele freqeuncy, genotype completeness and Hardy-Weinberg equilibrium p value are shown for each collection.



Supplementary Figure 2: Concordance, as a measure of accuracy, versus missing data. A set of 50,000 genome wide SNPs which were genotyped on both the Affymetrix 6.0 chip and Illumina 1.2M chip were used to calculate concordance. The Illumina data was called with Illuminus at the recommended confidence threshold of 0.95 and subject to stringent QC. The corresponding Birdsuite (blue) and Chiamo (red) genotypes were compared to Illuminus calls with increasing confidence thresholds, resulting in improved concordance but increased missing data. Concordance versus missing data for the **(a)** 1958BC collection and **(b)** UKBS. The dashed line shows the concordance and missing data rate when the recommended confidence threshold is used, 0.9 for Chiamo and 0.1 for Birdsuite.



Supplementary Figure 3: The percentage of SNPs removed using the information criteria, from specified minor allele frequency (MAF) and call rate bins where red → white corresponds to 0 → 100%.



Supplementary Figure 4: Q-Q plots for control-control (left) and case-control (right) trend association tests. Shaded areas indicate 95% confidence intervals on the expected distribution.